INTRODUCTION TO MACHINE LEARNING

DR. SUNEESH JACOB

FACULTY HANDLING EIE417

SEEE, SASTRA UNIVERSITY

CONDITIONAL PROBABILITY AND BAYES' THEOREM

- $P(B|A) \rightarrow$ The probability of occurrence of event B given the event A has already occurred.

- $P(A \cap B) \rightarrow$ The probability that the events A and B had occurred.

- Bayes' Theorem Proof:

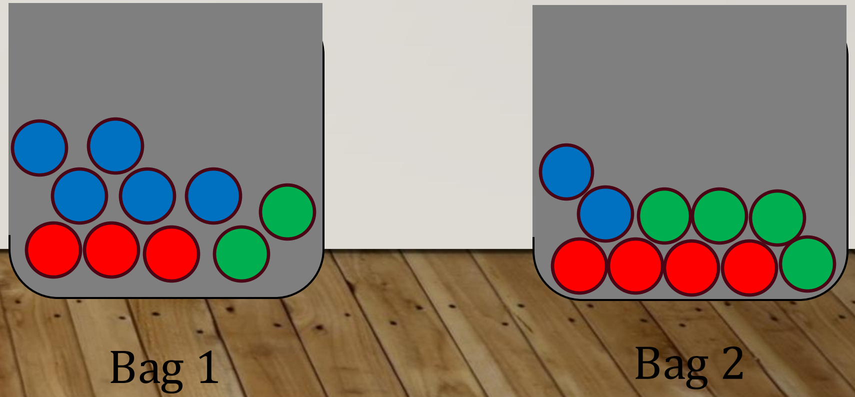
$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

$$\Rightarrow P(B|A) = \frac{P(B)P(A|B)}{P(A)}$$

EXAMPLE

- A bag is chosen randomly and one ball is drawn. It is found to be blue. What is the probability that it is from (a) Bag I and (b) Bag 2?
- $P(\text{Bag} = \text{bag } 1 | \text{Ball} = \text{blue})$
- $P(\text{Bag} = \text{bag } 2 | \text{Ball} = \text{blue})$

- $P(\text{Ball} = \text{blue} | \text{Bag} = \text{bag } 1) = \frac{1}{2}$
- $P(\text{Ball} = \text{blue} | \text{Bag} = \text{bag } 2) = \frac{1}{5}$
- $P(\text{Bag} = \text{bag } 1) = P(\text{Bag} = \text{bag } 2) = \frac{1}{2}$

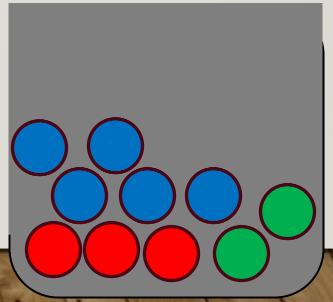


EXAMPLE

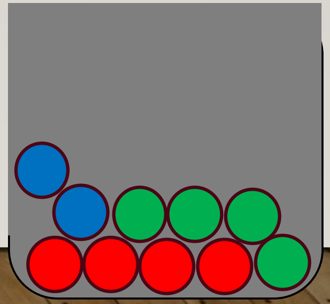
- A bag is chosen randomly and one ball is drawn. It is found to be blue. What is the probability that it is from (a) Bag 1 and (b) Bag 2?

- $P(\text{Ball} = \text{blue}) = P(\text{Bag} = \text{bag 1}) \cdot P(\text{Ball} = \text{blue}|\text{Bag} = \text{bag 1}) + P(\text{Bag} = \text{bag 2}) \cdot P(\text{Ball} = \text{blue}|\text{Bag} = \text{bag 2})$

$$\Rightarrow P(\text{Ball} = \text{blue}) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{5} = 0.35$$



Bag 1



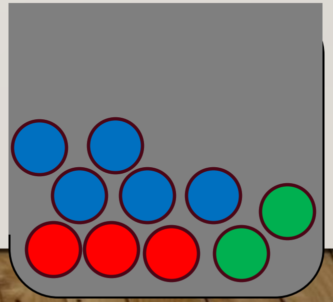
Bag 2

EXAMPLE

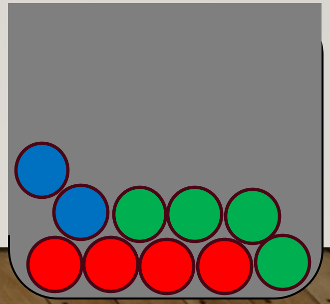
- A bag is chosen randomly and one ball is drawn. It is found to be blue. What is the probability that it is from (a) Bag 1 and (b) Bag 2?

- $P(\text{Bag} = \text{bag 1}|\text{Ball} = \text{blue}) = \frac{P(\text{Ball=blue}|\text{Bag=bag 1})P(\text{Bag=bag 1})}{P(\text{Ball=blue})} = \frac{\frac{1}{2}\times\frac{1}{2}}{0.35} = 0.714$

- $P(\text{Bag} = \text{bag 2}|\text{Ball} = \text{blue}) = \frac{P(\text{Ball=blue}|\text{Bag=bag 2})P(\text{Bag=bag 2})}{P(\text{Ball=blue})} = \frac{\frac{1}{5}\times\frac{1}{2}}{0.35} = 0.286$



Bag 1



Bag 2

DATA FOR AN ML PROBLEM

S. No.	Outlook	Temperature	Humidity	Wind	Match played
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rainy	Mild	High	Weak	Yes
5	Rainy	Cool	Normal	Weak	Yes
6	Rainy	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rainy	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rainy	Mild	High	Strong	No

THE ML PROBLEM

- Suppose

$$\left.\begin{array}{l}\text{Outlook is Sunny}\\ \text{Temperature is Cool}\\ \text{Humidity is High}\\ \text{Wind is Strong}\end{array}\right\} x$$

What is the prediction of the match?

SOLUTION (NAÏVE BAYES CLASSIFIER)

- To be found:
  - $P(\text{Match} = \text{Yes}|x)$
  - $P(\text{Match} = \text{No}|x)$
- From the table:
  - $P(\text{Outlook} = \text{Sunny}|\text{Match} = \text{Yes}) = \frac{2}{9}$
  - $P(\text{Temperature} = \text{Cool}|\text{Match} = \text{Yes}) = \frac{3}{9}$
  - $P(\text{Humidity} = \text{High}|\text{Match} = \text{Yes}) = \frac{3}{9}$
  - $P(\text{Wind} = \text{Strong}|\text{Match} = \text{Yes}) = \frac{3}{9}$

  - $P(\text{Outlook} = \text{Sunny}|\text{Match} = \text{No}) = \frac{3}{5}$
  - $P(\text{Temperature} = \text{Cool}|\text{Match} = \text{No}) = \frac{1}{5}$
  - $P(\text{Humidity} = \text{High}|\text{Match} = \text{No}) = \frac{4}{5}$
  - $P(\text{Wind} = \text{Strong}|\text{Match} = \text{No}) = \frac{3}{5}$

From the table:

$$P(\text{Match} = \text{Yes}) = \frac{9}{14}$$

$$P(\text{Match} = \text{No}) = \frac{5}{14}$$

SOLUTION

- $P(x|\text{Match} = \text{Yes}) = P((\text{Outlook} = \text{Sunny}, \text{Temperature} = \text{Cool}, \text{Humidity} = \text{High}, \text{Wind} = \text{Strong}) \mid \text{Match} = \text{Yes})$
- $P(x|\text{Match} = \text{No}) = P((\text{Outlook} = \text{Sunny}, \text{Temperature} = \text{Cool}, \text{Humidity} = \text{High}, \text{Wind} = \text{Strong}) \mid \text{Match} = \text{No})$

- If the input parameters are assumed to be independent of each other (i.e., Naïve), then
- $P(x|\text{Match} = \text{Yes}) = P(\text{Outlook} = \text{Sunny}|\text{Match} = \text{Yes}) \cdot P(\text{Temperature} = \text{Cool}|\text{Match} = \text{Yes}) \cdot P(\text{Humidity} = \text{High}|\text{Match} = \text{Yes}) \cdot P(\text{Wind} = \text{Strong}|\text{Match} = \text{Yes})$
- $P(x|\text{Match} = \text{No}) = P(\text{Outlook} = \text{Sunny}|\text{Match} = \text{No}) \cdot P(\text{Temperature} = \text{Cool}|\text{Match} = \text{No}) \cdot P(\text{Humidity} = \text{High}|\text{Match} = \text{No}) \cdot P(\text{Wind} = \text{Strong}|\text{Match} = \text{No})$
- $P(x|\text{Match} = \text{Yes}) = \frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}$
- $P(x|\text{Match} = \text{No}) = \frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}$
- $P(\text{Match} = \text{Yes}) = \frac{9}{14}$
- $P(\text{Match} = \text{No}) = \frac{5}{14}$

SOLUTION

- $P(\text{Match} = \text{Yes}|x) = \frac{P(x|\text{Match=Yes}) \cdot P(\text{Match=Yes})}{P(x)} = \frac{\left(\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9}\right)\left(\frac{9}{14}\right)}{P(x)} = \frac{0.0053}{P(x)}$

- $P(\text{Match} = \text{No}|x) = \frac{P(x|\text{Match=No}) \cdot P(\text{Match=No})}{P(x)} = \frac{\left(\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5}\right)\left(\frac{5}{14}\right)}{P(x)} = \frac{0.0206}{P(x)}$

- We know that

$$P(\text{Match} = \text{Yes}|x) + P(\text{Match} = \text{No}|x) = 1$$

$$\Rightarrow \frac{0.0053}{P(x)} + \frac{0.0206}{P(x)} = 1$$

$$\Rightarrow P(x) = 0.0053 + 0.0206 = 0.0259$$

- Hence
  - $P(\text{Match} = \text{Yes}|x) = \frac{0.0053}{P(x)} = 0.2046$
  - $P(\text{Match} = \text{No}|x) = \frac{0.0206}{P(x)} = 0.7954$

GAUSSIAN NAÏVE BAYES CLASSIFIER

S. No.	Temperature (°C)	Humidity (%)	Windspeed (km/h)	Rain Tomorrow
1	30	70	10	No
2	22	85	5	Yes
3	25	60	15	No
4	18	90	7	Yes
5	26	65	12	No
6	28	75	9	No
7	24	80	14	Yes
8	21	95	6	Yes
9	29	50	11	No
10	23	55	8	No

GAUSSIAN NAÏVE BAYES CLASSIFIER

- To find:
  - Rain prediction for tomorrow, given
    - Temperature = $27°C$
    - Humidity = 80%
    - Windspeed = $10\ kmph$

    $\left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} x$
  - $P(\text{Rain} = \text{No}|x) = ?$
  - $P(\text{Rain} = \text{Yes}|x) = ?$

  - $P(\text{Rain} = \text{No}|x) = \frac{P(x|\text{Rain} = \text{No})P(\text{No})}{P(x)}$
  - $P(\text{Rain} = \text{Yes}|x) = \frac{P(x|\text{Rain} = \text{Yes})P(\text{Yes})}{P(x)}$

GAUSSIAN NAÏVE BAYES CLASSIFIER

- $\mu_{\text{T,No}} = \frac{30+25+26+28+29+23}{6} = 26.833$
- $\mu_{\text{H,No}} = \frac{70+60+65+75+50+55}{6} = 62.5$
- $\mu_{\text{W,No}} = \frac{10+15+12+9+11+8}{6} = 10.833$

- $\sigma_{\text{T,No}} = \sqrt{\frac{(30-26.833)^2+(25-26.833)^2+(26-26.833)^2+(28-26.833)^2+(29-26.833)^2+(23-26.833)^2}{6}} = \sqrt{5.806}$
- $\sigma_{\text{T,No}} = \sqrt{\frac{(70-62.5)^2+(60-62.5)^2+(65-62.5)^2+(75-62.5)^2+(50-62.5)^2+(55-62.5)^2}{6}} = \sqrt{82.5}$
- $\sigma_{\text{W,No}} = \sqrt{\frac{(10-10.833)^2+(15-10.833)^2+(12-10.833)^2+(9-10.833)^2+(11-10.833)^2+(8-10.833)^2}{6}} = \sqrt{5.136}$

GAUSSIAN NAÏVE BAYES CLASSIFIER

- $\mu_{\text{T,Yes}} = \frac{22+18+24+21}{4} = 21.25$
- $\mu_{\text{H,Yes}} = \frac{85+90+80+95}{4} = 87.5$
- $\mu_{\text{W,Yes}} = \frac{5+7+14+6}{4} = 8$
- $\sigma_{\text{T,Yes}} = \sqrt{\frac{(22-21.25)^2+(18-21.25)^2+(24-21.25)^2+(21-21.25)^2}{4}} = \sqrt{5.188}$
- $\sigma_{\text{T,Yes}} = \sqrt{\frac{(85-87.5)^2+(90-87.5)^2+(80-87.5)^2+(95-87.5)^2}{4}} = \sqrt{31.25}$
- $\sigma_{\text{W,Yes}} = \sqrt{\frac{(5-8)^2+(7-8)^2+(14-8)^2+(6-8)^2}{4}} = \sqrt{11.5}$

GAUSSIAN NAÏVE BAYES CLASSIFIER

$$P(T|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{T,No}} e^{-\frac{1}{2}\left(\frac{T-\mu_{T,No}}{\sigma_{T,No}}\right)^2}$$

$$P(H|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{H,No}} e^{-\frac{1}{2}\left(\frac{H-\mu_{H,No}}{\sigma_{H,No}}\right)^2}$$

$$P(W|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{W,No}} e^{-\frac{1}{2}\left(\frac{W-\mu_{W,No}}{\sigma_{W,No}}\right)^2}$$

$$P(T = 27°C|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{T,No}} e^{-\frac{1}{2}\left(\frac{27-\mu_{T,No}}{\sigma_{T,No}}\right)^2} = 0.165$$

$$P(H = 80\%|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{H,No}} e^{-\frac{1}{2}\left(\frac{80-\mu_{T,No}}{\sigma_{T,No}}\right)^2} = 0.017$$

$$P(W = 10kmph|\text{Rain} = \text{No}) = \frac{1}{\sqrt{2\pi}\sigma_{W,No}} e^{-\frac{1}{2}\left(\frac{10-\mu_{T,No}}{\sigma_{T,No}}\right)^2} = 0.172$$

GAUSSIAN NAÏVE BAYES CLASSIFIER

$$P(T|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{T,Yes}} e^{-\frac{1}{2}\left(\frac{T-\mu_{T,Yes}}{\sigma_{T,Yes}}\right)^2}$$

$$P(H|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{H,Yes}} e^{-\frac{1}{2}\left(\frac{H-\mu_{H,Yes}}{\sigma_{H,Yes}}\right)^2}$$

$$P(W|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{W,Yes}} e^{-\frac{1}{2}\left(\frac{W-\mu_{W,Yes}}{\sigma_{W,Yes}}\right)^2}$$

$$P(T = 27°C|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{T,Yes}} e^{-\frac{1}{2}\left(\frac{27-\mu_{T,Yes}}{\sigma_{T,Yes}}\right)^2} = 0.018$$

$$P(H = 80\%|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{H,Yes}} e^{-\frac{1}{2}\left(\frac{80-\mu_{T,Yes}}{\sigma_{T,Yes}}\right)^2} = 0.066$$

$$P(W = 10kmph|\text{Rain} = \text{Yes}) = \frac{1}{\sqrt{2\pi}\sigma_{W,Yes}} e^{-\frac{1}{2}\left(\frac{10-\mu_{T,Yes}}{\sigma_{T,Yes}}\right)^2} = 0.114$$

GAUSSIAN NAÏVE BAYES CLASSIFIER

$$P(\text{Rain} = \text{Yes}) = \frac{4}{10} = 0.4$$

$$P(\text{Rain} = \text{No}) = \frac{6}{10} = 0.6$$

GAUSSIAN NAÏVE BAYES CLASSIFIER

- To find:
  - Rain prediction for tomorrow, given
    - Temperature = $27°C$
    - Humidity = 80%
    - Windspeed = $10\ kmph$

$$\left.\begin{array}{l}\text{Temperature} = 27°C \\ \text{Humidity} = 80\% \\ \text{Windspeed} = 10\ kmph\end{array}\right\} x$$

- $P(x|\text{Rain} = \text{No}) = P((T = 27°C \cap H = 80\% \cap W = 10kmph)|\text{Rain} = \text{No})$

$= P(T = 27°C|\text{Rain} = \text{No}) \cdot P(H = 80\%|\text{Rain} = \text{No}) \cdot P(W = 10kmph|\text{Rain} = \text{No}) = 0.165 \times 0.017 \times 0.172 = 0.000482$

- $P(x|\text{Rain} = \text{Yes}) = P((T = 27°C \cap H = 80\% \cap W = 10kmph)|\text{Rain} = \text{Yes})$

$= P(T = 27°C|\text{Rain} = \text{Yes}) \cdot P(H = 80\%|\text{Rain} = \text{Yes}) \cdot P(W = 10kmph|\text{Rain} = \text{Yes}) = 0.018 \times 0.066 \times 0.114 = 0.000135$

GAUSSIAN NAÏVE BAYES CLASSIFIER

- $P(\text{Rain} = \text{No}|x) = \frac{P(x|\text{Rain} = \text{No})P(\text{No})}{P(x)} = \frac{0.000482 \times 0.6}{P(x)} = \frac{0.00029}{P(x)}$
- $P(\text{Rain} = \text{Yes}|x) = \frac{P(x|\text{Rain} = \text{Yes})P(\text{Yes})}{P(x)} = \frac{0.000135 \times 0.4}{P(x)} = \frac{0.000054}{P(x)}$

- We know that $P(\text{Rain} = \text{No}|x) + P(\text{Rain} = \text{Yes}|x) = 1$

$$\Rightarrow \frac{0.00029}{P(x)} + \frac{0.000054}{P(x)} = 1$$

$$\Rightarrow P(x) = 0.0003447$$

GAUSSIAN NAÏVE BAYES CLASSIFIER

- $P(\text{Rain} = \text{No}|x) = \frac{0.00029}{P(x)} = 0.8435$
- $P(\text{Rain} = \text{Yes}|x) = \frac{0.000054}{P(x)} = 0.1565$

- Since $P(\text{Rain} = \text{No}|x) > P(\text{Rain} = \text{Yes}|x)$, the prediction for the input x is that the rain will not fall tomorrow.